AD-A047 838 ARMY TROPIC TEST CENTER APO NEW YORK 09827 F/G 5/5
A NEW APPROACH TOWARD OBTAINING QUANTIFIED SUBJECTIVE TEST DATA--ETC(U)
MAY 77 R L WILLIAMSON, D A DOBBINS
UNCLASSIFIED USATTC-7705001 NL

1 OF 1
ADA047838

END
DATE
FILMED
1-78
DDC

AD____________________

TECOM PROJECT NO. 7 CO IL7 TT1 001

USATTC REPORT NO. 7705001

A NEW APPROACH TOWARD OBTAINING QUANTIFIED SUBJECTIVE TEST DATA

By

R. L. Williamson
D. A. Dobbins

MAY 1977

Approved for public release; distribution unlimited.

UNITED STATES ARMY TROPIC TEST CENTER

FORT CLAYTON, CANAL ZONE

DISPOSITION INSTRUCTIONS*

Destroy this report when no longer needed. Do not return it to the originator.

DISCLAIMER

The findings in this report are not to be construed as an official Department of the Army position, unless so designated by other authorized documents. The use of trade names in this report does not constitute an official endorsement or approval of the use of such commercial hardware or software. This report may not be cited for purposes of advertisement.

* These disposition instructions do not apply to the record copy (AR 340-18).

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER
TECOM PROJECT NO. 7 CO IL7 TT1 001

2. GOVT ACCESSION NO.

3. RECIPIENT'S CATALOG NUMBER

4. TITLE (and Subtitle)
6 A NEW APPROACH TOWARD OBTAINING QUANTIFIED SUBJECTIVE TEST DATA.

5. TYPE OF REPORT & PERIOD COVERED
9 FINAL REP[redacted]T.

6. PERFORMING ORG. REPORT NUMBER
14 USATTC [redacted]-7705[redacted]1

7. AUTHOR(s)
10 R. L. WILLIAMSON
D. A. DOBBINS

8. CONTRACT OR GRANT NUMBER(s)

9. PERFORMING ORGANIZATION NAME AND ADDRESS
US ARMY TROPIC TEST CENTER
ATTN: STETC-TD
APO NEW YORK 09827

10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
16 1T061101A91A

11. CONTROLLING OFFICE NAME AND ADDRESS
US ARMY TEST AND EVALUATION COMMAND
ATTN: AMSTE-ME
ABERDEEN PROVING GROUND, MD 21005

12. REPORT DATE
11 MAY [redacted]77

13. NUMBER OF PAGES
12 21 p.

14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)

15. SECURITY CLASS. (of this report)
UNCLASSIFIED

15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Effectiveness	Perception	Subjective data
Feeling	Psychophysical	Taxonomy
Logarithmic function	Questionnaire	Tropic Test Center
Measurement	Ratio Scale	
Opinion	Sensation	

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

This report proposes a method for obtaining ratio-scaled response data during subjective questioning of soldiers whose opinions are solicited during Army materiel tests. The approach is an adaptation of psychophysical measurement methods developed in recent years.

Types of measurement scales are reviewed; a tentative taxonomy of psychophysical terminology is proposed. Methods and instrumentation for selecting and validating response modes are outlined. Plans for using resultant subjective measures of effectiveness in parametric models for statistical inference are suggested.

642 290

DD FORM 1473, 1 JAN 73 — EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

TABLE OF CONTENTS

APPENDICES

PREFACE

This report covers the beginning of a separate phase of human factors test methodology development of the US Army Tropic Test Center. While past efforts have concentrated on human performance in the humid tropics (i.e., vision, audition, portability/load carrying, land navigation ability, rifle-fire accuracy), this effort turns to the subjective domain of *materiel evaluation*. The work was supported by the US Army In-House Laboratory Independent Research (ILIR) Program.

blank

2

I. INTRODUCTION

In tests of new military hardware, the Army has traditionally placed high value on the acceptance and preferences of user personnel. Hence, the concepts of troop tests, service tests, force development and operational tests have evolved, as distinguished from engineering and developmental tests. The latter two yield quantitative indices of hardware performance functioning, reliability and maintainability. The former are often assessed by subjective methods—and it is widely recognized that the subjective evaluations are critical to the deployment of new hardware items. They add the important human factor which is independent of engineering data. At the same time, subjective measures arouse suspicion and uneasiness among many system evaluators. Subjective measures are prone to sources of error, to include biases of interviewer and interviewees, resistance to change, sheer disinterest of test participants and the classical errors of halo, horn, hello-goodbye, central tendency, acquiescent response sets, and many more. *Guilford, 1954*

The Army has adopted two general approaches to resolve the problem. The first approach is to improve data acquisition techniques in obtaining information from soldiers. This involves making the subjective techniques more *systematic*. Since the 1930's, much effort and many improvements have been made in the development of structured interview techniques, standardized questionnaire development, rating scales, panel evaluation, and checklists. However, there have been no true state-of-the-art advances since the 1940's when the "forced-choice" evaluation technique was developed for personnel assessment. The questionnaire and rating scale technology being used in the 1970's is substantially the same as that in use during World World II.

The second approach, which has great popular appeal, is to make the subjective evaluations more *objective*. That is, the human factor is approached from a quantitative viewpoint. Instrumented performance courses have been developed to measure factors such as speed, accuracy, completeness, and relevance for a great variety of military tasks. Physiological indicators such as heart rate, body temperature, and basal metabolism have also been widely used in performance assessment. However, hundreds of studies have shown that objective performance measures do not predict the subjective expressions of test participants. Rather, performance measurement has furnished an indispensable but *independent* measure of the human factor. The technology of subjective assessment has come to a virtual standstill.

Over the past 25 years, a successful small scale and low visibility program has begun to show promise in the area of subjective measurement. The work has been carried on in various university laboratories in the general area of "psychophysical scaling." The intent of the present study is to transfer this technology to Army materiel testing.

In the process of describing human performance demands of new complex systems, it is important to recognize the existence of performance problems, to identify their source, and to measure their magnitudes. By eliminating or reducing the magnitude of the problem, the overall efficiency of a system may be increased. But the first step in eliminating a problem is often a subjective report of its existence. An operator or controller of a new system may express difficulty in its "handling," but not be able to pinpoint the source of the problem or directly measure its magnitude. Typical subjective measurement scales produce "category-scaled" data with units that are ordinal at best (allowing rank-order comparisons, but not statements as to amount of difference or absolute level). *Stevens, 1975* Category-scaled data may be contrasted to data derived from more scientific methods of systematic measurements that produce "ratio-scaled" data (having an absolute zero and units that may be legitimately manipulated mathematically).

II. PROBLEM AND OBJECTIVE

This investigation attacks the problem of obtaining a quantitative subjective measure of effectiveness (SMOE) during developmental testing of Army materiel. Typical category scales of current subjective questionnaires provide ordinal data that can neither be manipulated mathematically beyond simple summation nor can be analyzed statistically beyond nonparametric tests of partitioned responses. This investigation will develop ratio-scaled procedures for obtaining subjective questionnaire responses. The scientific method may then be applied to subjective data. SMOE may be delineated more precisely.

III. BACKGROUND

Current guides for subjective questioning, questionnaire design, and data analysis include a wide variety of techniques. Each technique, although well used and very useful up to a point, has the same scaling problem in varying degrees (TECOM Pam 602-1).* Except for free-response subjective questioning, the problem is that respondents are forced to conform to preset scales, therefore losing the freedom to respond more sensitively and precisely according to their feelings and opinions. Free-answer or open-ended questioning is useful in exploratory studies where restrictions in response form may inhibit expression of potentially important personal insight, or is useful as a follow-up technique for amplifying or explaining scaled responses. In either case, resulting verbalizations are of more use in formulating questions than in documenting response levels; response scaling methods do not apply to the problem of this investigation. Questionnaire designs other than the open-ended type contain specific questions, each requiring a respondent to conform to a preset response mode. The most basic of these is the dichotomous mode where the response is the equivalent of *yes* or *no* (sometimes including a third *don't know* option). The constraint of the dichotomous response provides no sensitivity to the degree of "yesness" or "noness" that the respondent may be able to express. Although gestures or verbal comments may qualify these responses at the time they are recorded, analyses of the data are denied such advantages and are limited to the oversimplified response split.

The next level of response sophistication contains a host of categorically scaled mechanisms including multiple choice or checklist responses (where one or more of a number of alternative nominal categories are to be checked in preference to others), and rating scales (where the respondent is to select one category from an ordered series such as *no problem, very little problem, somewhat difficult, or very difficult*). Rating categories are verbal, numerical, or both. Some are composed of a numerical scale (rarely over 10 points on the continuum) combined with a verbal anchor at each of the extremes, but not at the middle points, such as:

1	2	3	4	5	6	7	8	9	10
Agree Strongly									Disagree Strongly

Although rating scales are used to gain a degree of sensitivity to variation in possible responses, and are superior to dichotomous scales in that respect, rating scales are nonetheless ordinal, with some (based on standardized phrases) achieving a quasi-interval nature. The points along rating scales are designed or assumed to be evenly spaced for purposes of analyzing response levels. However, unless the response categories of a rating scale have been shown to be equally spaced by way of standardized weighting procedures, then interval scaling cannot be assumed and the more powerful parametric statistics are inappropriate.

*TECOM Pamphlet 602-1, Vol. 1, Questionnaire and Interview Design (Subjective Testing Techniques), 25 July 1975.

A good example of the problem of this investigation may be seen by examining the nature of a type of subjective questioning that was left out of the foregoing paragraph for this purpose—the ranking question. A typical ranking question presents a number of items that the respondent is asked to put in rank order according to his personal preference. An example is:

Rank the following types of helmets in the order of your preference.

(1 = most preferred, 2 = next preferred, etc.)

(a) New, Type I ______

(b) New, Type II ______

(c) New, Type III ______

(d) Standard ______

A particular soldier who happened to like the new, Type I, helmet best would put the number 1 on the first line, followed perhaps by ranks of 4, 3, and 2 for the remaining items, respectively. Analyses of the data are limited to those appropriate for ordinal measurement because it cannot be inferred from the ranks, for instance, *how much more* that one helmet is preferred to the others. But suppose we lifted the restriction in range of assignable "ranks" from *1 through 4* to *0 through any number the soldier wanted to use*. Furthermore, because our society generally associates "largeness" with "goodness," we could turn the scale around and ask the soldier to think of zero as representing the least preferred helmet imaginable to him, and place no restriction on his assignment of a number to the helmet that is most preferred on his own scheme of preference.* Then the soldier whose preference for the new, Type *II*, helmet was extremely low, but not as low as something else that was not on the list, could rate new, Type II, as *9*, new, Type III, as *100*, the Standard as *210;* and, if new, Type I, were far out on his own preference scale, he could rate it as *9000* if that number represented the way his preference ran.† By transforming the rating scale into an unconstrained numerical field and allowing the subject to match numbers to his feelings, a higher order of scaling occurs. The scale not only goes from ordinal to interval (where we can say that the difference between new, Type II, and new, Type III, preference was 91 = 100 - 9), but also goes to ratio (where a real zero preference level allows us to say that his preference for the new, Type I, helmet was 1000 times greater than for new, Type II, and 43 times greater than the Standard helmet).†

The foregoing illustrated the problems associated with category scaling of subjective questionnaire responses. The last example introduced the psychophysical measurement approach—cross-modality matching—that this investigation will take toward their solution.

IV. APPROACH

The general approach to the problem is to apply the ratio scaling techniques of psychophysical cross-modality matching to subjective questioning. Ratio scaling is a fairly recent state-of-the-art advancement in psychological measurement. For over 200 years, psychologists and physicists have been building a case that the intensity of a stimulus and

*The exact wording of the question and the basis of his preference (weight, shape, balance) would be important issues to resolve in an actual test, but need not be addressed here in a discussion of response scaling.

†Again, problems surrounding question wording, practice, varying ranges among individuals, and data reduction techniques are areas of needed research, but the scaling concept may be discussed separately from their solution at this time.

the magnitude of the sensation were related according to a logarithmic function. However, in the last 25 years S. S. Stevens, among others, showed that the logarithmic function is biased for most continua because data were collected through partitioning procedures–judging subdivisions or apparent differences through such methods as just noticeable differences (JNDs), rather than judging ratios.

> "On prothetic continua–those continua concerned with intensity or amount, the variability–and hence the JND, tends to increase in proportion to the magnitude. Consequently, the counting off of prothetic JND leads to a logarithmic function. When partitioning procedures (including bisection and category scaling) are applied to prothetic continua, there results a biased function, a function that is curved relative to the scale of magnitude determined by ratio scaling procedures.
>
> "Scales of perceptual magnitude may be created by asking observers to match numbers to stimuli. Beginning in 1953, it was shown that on prothetic continua the perceived magnitude increases as a power function of the stimulus magnitude. Each modality has its own exponent, although the value of the exponent may change with adaptation, contrast and other parameters of our experiment. The exponent of the power function determines the curvature of the function. The basic principle that underlies the power law is that equal stimulus ratios produce equal sensation ratios." *Stevens, 1975, pp 35-36*

So, the useful basic concept behind psychophysical scaling is that "measurement" is a process or procedure that can be applied to sensations, perceptions, or subjective questionnaire responses. Measurement is much broader than counting or enumerating things in terms of a physically countable unit. Stevens, 1975, proposed "matching" to be the basis of all measurement; counting was explained as a special case of matching, where words and numerals have come to be substituted for the original procedure of matching pebbles, notches on a stick, or tallies to the items of interest (measuring numerosity). Stevens regarded measurement as a "two-part endeavor, consisting on the one hand of manipulations and on the other of models." He explained the measurement procedure as a "schemapiric enterprise . . . the schematics of mathematics and the empirics of laboratory operations. Mathematics can mirror manipulations, but it no longer legislates their freedom. We now recognize that measurement extends to wherever . . . we can invent systematic rules for pinning numbers on things." When the rules involve a procedure for directly matching a perceived magnitude along one continuum to a perceived magnitude on another continuum, the magnitude of the sensation has been shown to be a power function of the stimulus, and a ratio-scaled measurement results. *Stevens, 1975*

In order to understand the nature of the problem addressed by this investigation, and the scope of the methodology stated in the next section, the following measurement and psychophysical terms and relationships are offered. They have been compiled from a review of approximately 200 articles and books published in the area of psychophysics in the last five years. Terms and examples not referenced upon their initial use are the authors'. In an attempt to bring the cascade of terms into some perspective, they have been placed (forced in some cases, perhaps beyond the limits of their original intent) into a tentative taxonomy that will undoubtedly change as investigations progress. The terms are presented first; the taxonomy follows.

Types of Measurement Scales or Continua

Prothetic *Stevens, 1975*: Refers to quantitative continua on which the degree of a stimulus or response may be scaled. The stimulus-response (SR) is additive, allowing a measurement of "how much" or "how much more" a stimulus is presented or a response is made. Contrasted to metathetic.

Metathetic *Stevens, 1975*: Refers to qualitative, positional continua on which different kinds of sensations may be categorized. Positions on the continuum are independent, allowing substitutive measurement of "different from in kind." Stevens gives an example that "... sweet is (metathetically) different from sour, although both may vary (prothetically) from strong to weak."

Heterothetic: Refers to an SR relationship wherein both prothetic and metathetic continua must be measured for its description.

Interoceptive *Sullivan, 1973*: Refers to subjective, judgmental aspects of a sensation for which no direct physical measurement is appropriate for all individuals. May be metathetically scaled to distinguish basic properties. May be prothetically scaled to distinguish among levels of intensity. Examples are anxiety, hunger, anger, thirst, and fatigue.

Exteroceptive *Sullivan, 1973*: Refers to objective, physical aspects of a sensation for which reliable data may be fixed to the stimulus. Is prothetically scaled. Examples are brightness, loudness, heat, and weight.

Heteroceptive: Refers to an SR relationship wherein both interoceptive and exteroceptive continua are necessary for its description.

Intensive: Designates investigation of a single (one) stimulus or response that may be either simple (consisting of one prothetically measured part) or complex (consisting of a set of more than one interrelated, prothetically measured component parts); contrasted to "extensive." A suggested example of a simple intensive stimulus is a straight line measured by its length. An example of a complex intensive stimulus may be brightness, as measured by flash duration and luminance. *Marks, 1974* Prothetic measures must be stated in order to ensure that component parts are totally accounted for and can be interrelated. Otherwise, the existence of an undiscovered unrelated part may require a redesignation to "extensive." Designations of intensive are based on "current knowledge" and are therefore tentative at best.

Extensive: Designates investigation of compound (two or more unrelated) stimuli or responses, each of which may be either simple or complex. An example is the size, operability, portability, maintainability, and safety of a weapon.

Tentative Taxonomy Based on Selected Psychophysical Terminology

Type of Measurement	Scale Continua	Nature of Stimulus (S)	Nature of Response (R)	Research Product
Psychophysical (Stevens, 1975)	Intensive heteroceptive prothetic	Single simple physical (measured)	Single simple sensations or perceptions (measured)	Relationship between S & R (power function exponent)

Example: Stevens[1975] provides many examples of power function exponents (β), where the R magnitude (ψ) grows as a power function of the S magnitude (ϕ) in the form $\psi = K\phi^{\beta}$, where K is a constant dependent on the units of measurement. For perceived loudness of a 3000Hz tone, $\beta = 0.67$; for discomfort from whole body irradiation, $\beta = 0.7$; for perceived heaviness of lifted weight, $\beta = 1.45$.

☆ ☆ ☆

Type of Measurement	Scale Continua	Nature of Stimulus (S)	Nature of Response (R)	Research Product
Psychosensory (Marks, 1974)	Intensive interoceptive prothetic	Physical (measurement irrelevant)	Single complex sensations (measured)	Relationship among R components (equation valid for all levels of S)

Example: Marks[1974] used the example that the loudness of a sound heard by two ears (L_b) equals the sum of the loudness heard by the left (L_l) and right (L_r) ears ($L_b = L_l + L_r$). The magnitude of the stimulus is irrelevant and need not be measured to determine the psychosensory function.

☆ ☆ ☆

Type of Measurement	Scale Continua	Nature of Stimulus (S)	Nature of Response (R)	Research Product
Sensory-physical (Marks, 1974)	Intensive exteroceptive prothetic	Single complex physical (measured)	Sensations or perceptions (measurement irrelevant)	Relationship among S components (equation valid for all levels of R)

Example: Marks[1974] also provided the example of Bloch's law of temporal summation wherefrom constant brightness (K_b) is the product of flash duration (t) and luminance (L); i.e., ($L \times t = K_b$). The magnitude of the response level on a scale of brightness is irrelevant and need not be measured by psychophysical methods to determine the sensory-physical relationship.

☆ ☆ ☆

Type of Measurement	Scale Continua	Nature of Stimulus (S)	Nature of Response (R)	Research Product
Psycho-attitudinal	Extensive heteroceptive heterothetic	Compound complex physical and situational (measurement irrelevant)	Compound simple attitudes (measured)	Collective evaluation from diverse separate R elements (summary & analysis of R pattern)

(cont)

Example: Evaluating various aspects of attitude of a population toward an object or situation such as assessing troop acceptability of a new military item used in a harsh environment. Psychophysical measurement methods used to obtain multiple ratio-scaled questionnaire Rs. Ratio-scaling provides level and pattern of attitudes toward object or situation; provides base for comparisons across objects or situations.

☆ ☆ ☆

The preceding taxonomy separates types of measurement on the bases of the degree of complexity and duplicity of the stimulus and response, and whether or not the stimulus and response must be measured. For *psychophysical* measurement, both the stimulus and response must be measured; each is single and simple; measurement relates the magnitude of one stimulus to the magnitude of one response. For *psychosensory* measurement, only the response is measured; the response is single but complex; measurement relates the magnitudes of the component parts of the response. For *sensory-physical* measurement, only the stimulus is measured; the stimulus is single but complex; measurement relates the magnitudes of the component parts of the stimulus. For *psycho-attitudinal* measurement, only the responses are measured; the responses are compound and simple; measurement describes the magnitudes of diverse responses that may or may not be related. It is within the final or psycho-attitudinal type of measurement of the preceding taxonomy that the current investigation lies. The reason is that the goal of this investigation is to develop a subjective measure of effectiveness (and associated instrumentation) that would require a single procedure for measuring a *series* of attitudes, the natures of which may be quite different (compound attitudes). The other three types of measurement are aimed at intensive investigations of single sensations or perceptions, measured by highly specialized procedures and instrumentation that may be of little value across the many responses to a questionnaire.

V. METHOD

SUBJECTS

During initial SMOE developmental stages of each of the response modes, inhouse personnel will be used on the basis of their availability. During validation stages, random samples of in-house personnel will be used in addition to representative troops from the 193d Infantry Brigade (Canal Zone). It has been found that a group of 10 individuals provides data stable enough for validating psychophysical power functions.[*Stevens, 1975, p 30*] During field trials, test subjects will be the personnel who operate, maintain, or are otherwise involved in active tropic testing of materiel from whom subjective questionnaire responses would normally be obtained.

PROCEDURE

General SMOE Development Program. To show how the specific procedures of this report fit into the longer term objectives of the SMOE development program, it will be helpful to outline the general program procedure first.

a. **Modal development** is the first stage. Each response mode investigated must undergo a modal development stage where materials and methods will be established, trial tested, and honed to a point where validation of modal response patterns may be attempted.

b. **Psychophysical validation** will be the second stage of the program. Many response modes have been used by psychophysical experimenters over the last 25 years. Each has an associated power function exponent that has been replicated many times. The established exponents, then, may serve as criteria against which the materials and procedures developed in the previous stage may be validated. For instance, it has been determined that when a person draws lines on a paper to represent the magnitudes of numbers spoken to him, the lengths of the lines are in a 1:1 ratio to the magnitude of the numbers he hears. Therefore, if producing a line were considered as a useful way of gauging the intensity of an attitude (or each of a series of attitudes as on a questionnaire), then the procedure for producing the line (exact instructions, size of paper) should be shown to yield a 1:1 relationship to the magnitude of spoken numbers as an initial *calibration* step. Similarly, if producing a matching tone were to be a basis for measurement, the calibration step would be to replicate the .67 power exponent found to exist in magnitude estimations of tones. Response modes that do not compare favorably to appropriate criterion values will then be recycled through the development stage as many times as necessary to insure that the procedures for obtaining subjective responses via the mode in use do, indeed, produce ratio-scaled responses with acceptable power function exponents. Figures 1 through 4 show various methods of ratio scaling techniques.

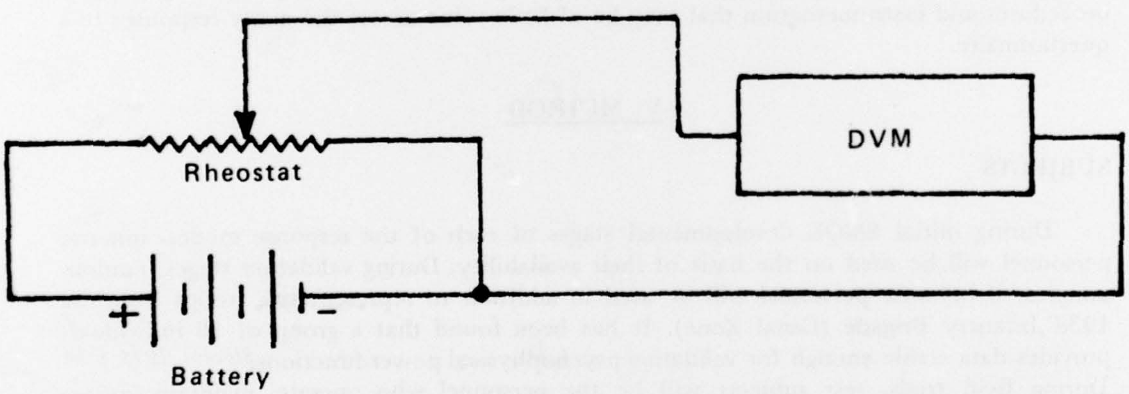

Figure 1. Ratio Scaling by Varying Voltage—A subject could control a voltage from some minimum to some maximum by changing the position of a rheostat. His response would be read as a number on a digital voltmeter.

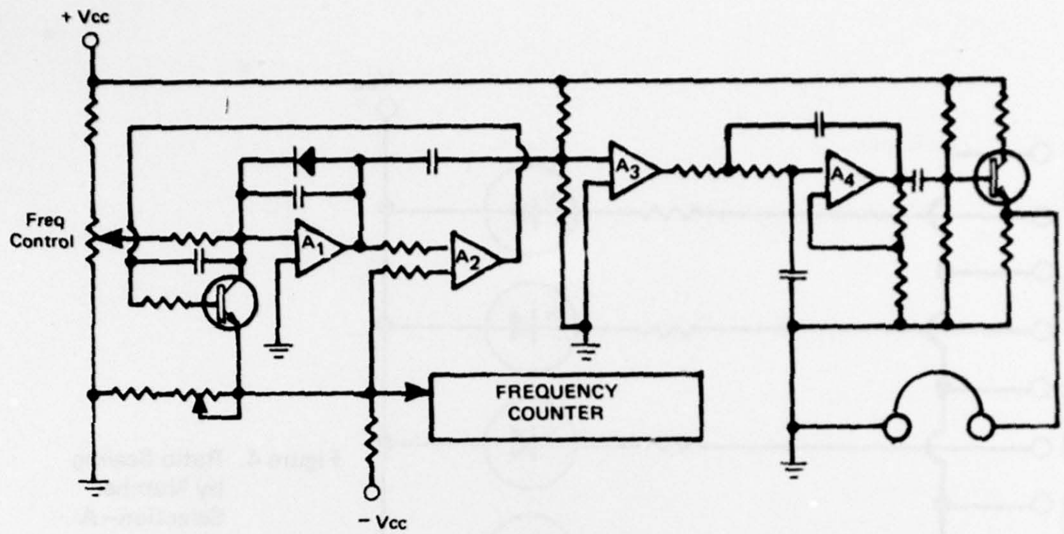

Figure 2. Ratio Scaling by Use of Frequency Control—A subject would control a variable frequency source. By listening with headphones he could set his response according to the highness or lowness of the frequency. His response would read as a number on a *frequency counter. The number could vary from zero to the limit* of the frequency source.

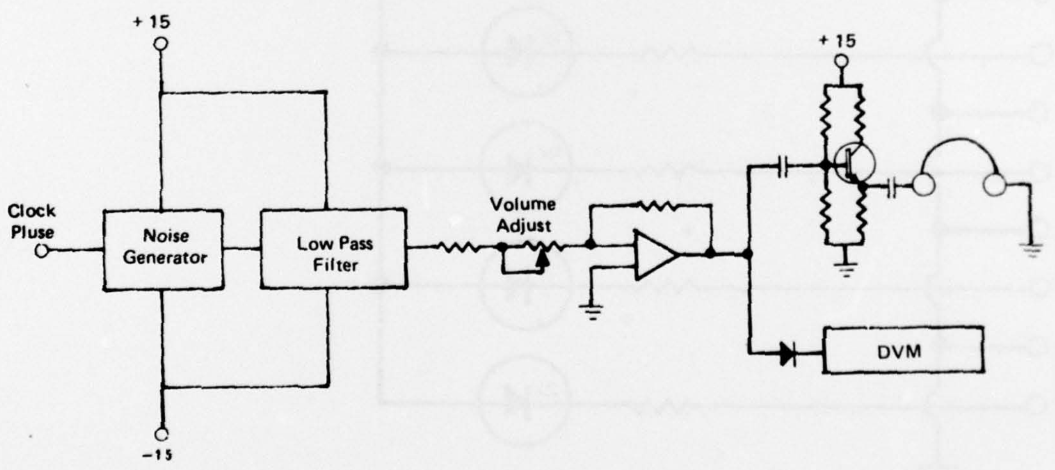

Figure 3. Ratio Scaling by Use of Loudness Control—A subject would move a volume control and set the loudness of a sound in accordance with his likes or dislikes. For example, the louder the sound the more he likes or dislikes an item. His response could be read as a number either on a voltmeter or a sound-level meter. *Stevens, 1975*

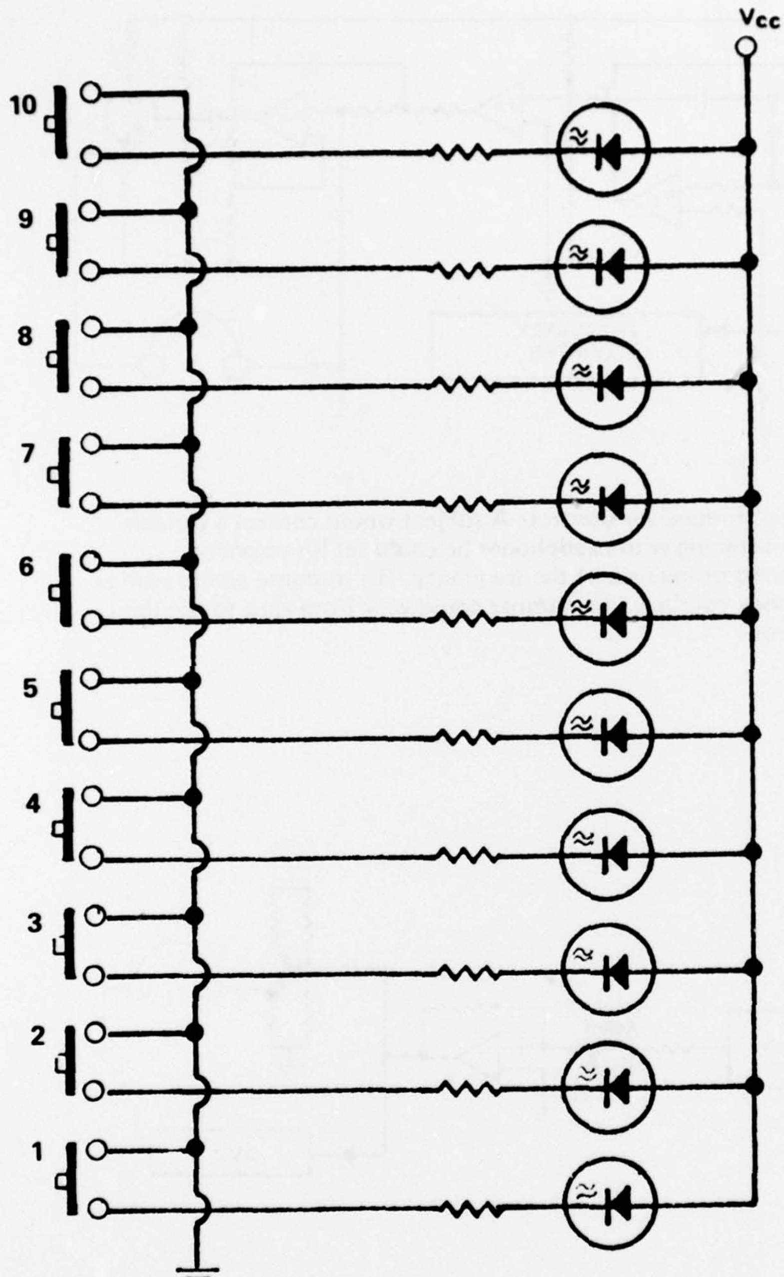

Figure 4. Ratio Scaling by Number Selection—A subject could indicate preference by selecting any number between 1 and 10 and observe his selection on a digital LED readout.

c. **Psycho-attitudinal validation** is scheduled to follow the psychophysical validation discussed above. Each response measurement mode will be tested for its validity to measure known amounts of difficulty to perform soldier-item tasks that are representative of subjectively measured human factors aspects of tropic testing. In order to ensure that all important human factors aspects are covered in this stage of the investigation, two mock test items will be built. The purpose will be to ensure that the amount of difficulty to perform tasks, the stimuli, can be controlled and measured to provide known criteria. Each mock item will be identical to the other, except for superficial aspects that make one appear to be a test item and the other to be a control item. The nature of the items may not be unlike a chemical-biological shelter system used in the tropics. Various soldier-item interfaces (such as force to open a door, weight of movable components, light levels, noise levels, control manipulation force, temperature, clarity of operating manuals) will be set at different known levels in each of the items (with the *difficult* interfaces not necessarily being all in the same item). At this stage of SMOE development, combat troops who would normally use such an item will be tested for their subjective responses to difficulty in performing the tasks. Ratio-scaled subjective responses obtained through the mode(s) being developed may be compared to known, preset, levels of difficulty in carefully designed and controlled experiments. Examples of types of validity and reliability studies that may be conducted are: ability of a psychophysical response mode to reflect various known levels of difficulty; sensitivity of a response mode to small differences in preset levels of difficulty (at low levels, intermediate levels, and high levels of difficulty); stability of response level when preset levels are identical in test and control items—tested same point in time; stability of response levels over varying lengths of time between trials on the same item, set to the same level each time; comparisons among various response modes, including typical category scales, on all of the above; and reaction of troops to measurement methods.

d. **Field validation** will be conducted after the various response modes have been validated and compared as outlined in the preceding paragraphs. The most suitable modes will be tested in the field during regularly scheduled tropic tests of materiel items. Comparison of combat troop response to test items will be made using data from ratio-scaled SMOE and data obtained from typical category-scaled techniques. An example of field validation would be a series of simple experiments using two items, standard and new; say entrenching tools—standard "old" and NARADCOM's "new." Have 20 soldiers dig two holes each; then, use a potentiometer to compare preferences. Also use one or two paper and pencil scales; analyze for: (1) reliability of ratio scaling from soldier to soldier; (2) correlation between ratio scaling and paper/pencil scales.

e. **SMOE modeling** will be performed with techniques that prove to be effective for obtaining ratio-scaled subjective responses for a variety of typical materiel items scheduled for tropic testing. Techniques will be formalized into standard test operation procedures and associated instrumentation suitable for use throughout the Army.

Program Application.

a. As an example of how the SMOE program would work, let us consider a typical situation in which a test item, say a new protective fragmentation vest or helmet where item acceptance relies heavily on subjective data from troops, is compared with a standard item.

The comparison is generally required in several environments (temperate, humid tropic, arctic, and desert), in numerous tactical situations (attack, defense, parachuting), and a myriad of functional capacities (body movement, stability, comfort, compatibility, vulnerability, maintainability safety, confidence). Each of the functional capacities may be covered by several specific questions on the degree of difficulty in performing specific tasks (moving the head, keeping balance, staying cool/warm, interfering with rifle firing, seeing, providing camouflage).

b. The test situation calls for a multivariate analysis that would not only uncover major problems with the test/control item, but also identify possible interaction effects; the test system may be of greater utility in one environment and of lesser utility in another environment, with the opposite being true for the control system. Given a coordinated test program where methods and instrumentation are standardized (e.g., potentiometer slide and taped instructions and questions), ratio-scaled subjective data could be analyzed, for instance, in a 4 (environments) x 3 (tactics) x 8 (functions) x 5 (tasks nested within each function)--a powerful analytical tool not legitimate for typical subjective data.

APPENDIX A. BIBLIOGRAPHY

1. Anderson, Norman H. (U. California, San Diego) *Cross task validation of functional measurements using judgments of total magnitude.* Journal of Experimental Psychology, 1974 (Feb), Vol. 102(2), 226-233. Journal Abstract.

2. Auerbach, Carl. (Yeshiva U.) *A note on cross-modality matching.* Quarterly Journal of Experimental Psychology, 1973 (Nov). Vol. 24(4), 492-495. Journal Abstract.

3. Banks, William P. (Pomona Coll.) *A new psychophysical ratio scaling technique: Random production.* Bulletin of Psychonomic Society, 1973 (Apr) Vol. 1(4), 273-275. Journal Abstract.

4. Banks, William P., and Hill, David K. (Pomona, Coll.) *The apparent magnitude of numbers scaled by random production.* Journal of Experimental Psychology, 1974 (Feb). Vol. 102(2), 353-376. Journal Abstract.

5. Birnbaum, Michael H. (Kansas State U.) *Using contextual efforts to derive psychophysical scales.* Perception and Psychophysics, 1947 (Feb). Vol. 15(1), pp. 89–96. Journal Abstract.

6. Birnbaum, Michael H. and Veit, Clarice T. (Kansas State U.) *Scale convergence as a criterion for rescaling: Information integration with difference, ratio, and averaging tasks.* Perception and Psychophysics, 1974 (Feb). Vol. 15(1), 7-15. Journal Abstract.

7. Borg, Gunnar. *Perceived exertion during walking: A psychophysical function with two additional constants.* Reports from the Institute of Applied Psychology, U. Stockton, 1973, No. 39, 7p. Journal Abstract.

8. Borg, Gunnar. *On the importance of range differences on psychophysical functions.* Reports from the Institute of Applied Psychology, U. Stockholm, 1972, No. 31, 6p. Journal Abstract.

9. Borg, Gunnar. *The basic "noise constant" in the psychophysical function of perceived exertion.* Reports from the Institute of Applied Psychology, U. Stockholm, 1972, No. 33, 8p. Journal Abstract.

10. Borg, Gunnar. (U. Stockhom, Inst. of Applied Psychology, Sweden) *A note on a "dispersion method" in psychological scaling.* Reports from the Institute of Applied Psychology, U. Stockholm, 1971 No. 17, 4p. Journal Abstract.

11. Borg, Gunnar, Edstron, Carl G., and Markland, Gustaf. (U. Stockholm, Institute of Applied Psychology, Sweden) *A new method to determine the exponent for perceived force in physical work.* Catalog of Selected Documents in Psychology, 1972 (Win), Vol. 2, 10-11.

12. Cliff, Norman. (U. Southern California) *Scaling.* Annual Review of Psychology, 1973, Vol. 24, 473-500.

13. Dercole, K. L. and Davenport, W. G. (U. British Colombia, Vancouver, Canada) *Social Psychophysics; Measurement of attitude toward violence.* Perceptual and Motor Skills, 1974 (Feb), Vol. 38(1), 135-145. Journal Abstract.

14. Edgerly, John W. and Gaither, Gerald. (U. Tennessee, Student Counseling Center) *Assessing attitude intensity among student factions. A behavioral model.* Journal of Human Relations, 1971, Vol. 19(3), 377-393.

15. Eisler, Harris and Montgomery, Henry. (U. Stockholm, Psychological Lab, Sweden) *Are theoretical and realizable ideal conditions in psychophysics: Magnitude and category scales and their relation.* Perception and Psychophysics, 1974 (Aug), Vol. 16(1), 157-168. Journal Abstract.

16. Fenker, Richard M. (Texas Christian U.) *New dimensions in psychophysics: A sure cure for methodological dyspepsia.* US Army Human Engineering Laboratories Technical Memorandum, 1971 (Dec), No. 25-71, 55-58.

17. Fenker, Richard M., et al. *The Psychological Dimensions of Camouflaged Imagery,* Technical Memorandum 26-75. US Army Human Engineering Laboratories, Aberdeen Proving Ground, Maryland, October, 1975.

18. Fidell, Sanford, Pearsons, Karl and Bennett, Ricarda. (Bolt, Beranek and Newman, Canoga Park, CA) *Prediction of aural detectability of noise signals.* Human Factors, 1974 (Aug), Vol. 16(4), 373-383. Journal Abstract.

19. Garman, Bernard S., et al. (City Coll., City U. New York) *Linear representation of temporal location and Steven's Law.* Memory and Cognition, 1973 (Apr), Vol. 1(2), 169-171. Journal Abstract.

20. Guilford, J. P., *Psychometric Methods,* McGraw Hill, New York, 1954.

21. Herrick, Robert M. (US Naval Air Development Center, Warminster, PA) *Psychophysical methodology: VI. Random method of limits.* Perception and Psychophysics, 1973 (Jun), Vol. 13(3), 548-554. Journal Abstract.

22. Jacobs, Diana E. and Galanter, Eugene. *Estimates of utility function parameters from signal detection experiments.* New York, N.Y.: Colombia University, Psychophysics Lab, TR PLR-32, 1974. 18p.

23. Jeffriss, Lloyd A. (U. Texas, Applied Research Lab.) *The logistic distribution as an approximation to the normal curve.* Journal of the Acoustical Society of America, 1973 (May), Vol. 53(5), 1296. Journal Abstract.

24. Kneppreth, Narwood P., et al. *Techniques for the assessment of worth,* Technical Papers 254, US Army Research Institute for the Behavioral and Social Sciences, Arlington, VA 22209, Aug 1974.

25. Kranity, Davis H.; Atkinson, Richard C.; Luce, R. Duncan and Suppes, Patrick (Eds). (U. Michigan) *Contemporary developments in mathematical psychology: II. Measurements, psychophysics and neural information processing.* San Francisco, CA: W. H. Freeman, 1974. XV, 468 p.114.

26. Luce, R. Duncan. (Inst for Advanced Study, Princeton, N.J.) *What sort of measurement is psychophysical measurement?* American Psychologist, 1972 (Feb), Vol. 27(2), 96-106. Author Abstract.

27. MacAdam, David L. (Eastman Kodak Company, Research Lab, Rochester, N.Y.) *Uniform color scales.* Journal of the Optical Society of America, 1974 (Dec), Vol. 64(12), 1691-1702. Journal Abstract.

28. Marks, Lawrence F. (Yale U.) *On scales of sensation: Prolegomena to any future psychophysics that will be able to come forth as science.* Perception and Psychophysics, 1974 (Oct), Vol. 16(2), 358-376. Journal Abstract.

29. McCullough, Maurice L. and Clarke, Michael J. (Preston Polytechnic, England) *Human response to whole-body vibration: An evaluation of current trends.* Human Factors, 1974 (Feb) Vol. 16(1), 78-86. Journal Abstract.

30. McFadden, Dennis and Skinner, Robert B. (U. Texas, Austin) *On the form of psychometric functions for taste.* Perception and Psychophysics, 1974 (Apr), Vol. 15(2), 379-382. Journal Abstract.

31. Moskowitz, Howard R.; Scharf, Bertram and Stevens, Joseph C. (Eds.). (US Army Natick Lab, MA) Sensation and Measurement: Papers in honor of S.S. Stevens. Dordrecht, Holland: Dr. Reidel, 1974: Xiii, 469p.

32. Parker, Scott and Schneider, Bruce. (Colombia U.) *Nonmetric scaling of loudness and pitch using similarity and difference estimates.* Perception and Psychophysics, 1974 (Apr) Vol. 15(2), 238-242. Journal Abstract.

33. Peck, Robert E. *Thymometry: A technique for measuring feelings.* Behavioral Neuropsychiatry, 1973-1974 (Oct), Vol. 5(7-12), 8-12. Journal Abstract.

34. Pressy, Alexander W. (U. Monitaba, Winnipeg, Canada) *Measuring the Ponzo illusion with the method of production.* Behavior Research Methods and Instrumentation, 1974 (Jul), Vol. 6(4), 424-426.

35. Samet, Michael G., *Subjective interpretation of reliability and accuracy scales for evaluating military intelligence.* Technical Paper 260, US Army Institute for the Behaviorial and Social Sciences, Arlington, Virginia 22209, January 1975.

36. Schufeldt, Coral Vance, *A new approach to pilot rating scales,* Thesis, United States Naval Post Graduate School, Monterey, California 93940, September 1971.

37. Stevens, James P. (U. Cincinnati) *Step-down analysis and simultaneous confidence intervals in Manova;* Multivariate Behavioral Research, 1973 (Jul). Vol. 8(3), 391-402. Journal Abstract.

38. Stevens, S. S., *Perceived Level of Noise by Mark VII and Decibels (E).* Journal of the Acoustical Society of America Vol. 51, No. 2, Part 2, 1972.

39. Stevens, S. S., *Psychophysics,* John Wiley & Sons, New York, 1975 (pg 49).

40. Stone, Leroy A. (U. North Dakota) *Multidimensional scaling of persons in groups: Progress report.* Perceptual and Motor Skills, 1972 (Aug) Vol. 35(1), 35-42. Journal Abstract.

41. Sullivan, Richard. (Inst. for Study and Control of Anxiety, Fresh Meadows, N.Y.) *Anxiety: A method for scaling its relative magnitude and aversiveness.* Journal of Abnormal Psychology, 1973 (Dec), Vol. 82(3), 483-490. Journal Abstract.

42. Svensson, Leif T. and Szczygiel, Kristoffer (U. Stockholm, Sweden). *A matching method of successive approximations and its instrumentations.* Behavior Research and Instrumentation, 1974 (Jan), Vol. 6(1), 13-18. Journal Abstract.

43. Teghtsoonian, Robert (Smith Coll.) *Range effects in psychophysical scaling and a revision of Steven's law.* American Journal of Psychology, 1973 (Mar) Vol. 86(1), 3-27. Journal Abstract.

44. Virsu, Veijo and Lehtio, P.K. (U. Helsinki, Finland) *A microphotometer for measuring luminance distributions on a CRT.* Behavioral Research Methods and Instrumentation, 1975 (Jan), Vol. 7(1), 29-33.

45. Walker, R. and Adelman, C. (U. East Anglia, Center for Applied Research in Education, Norwich, England) *Interaction analysis in informal classrooms: A critical comment on the Flanders system.* British Journal of Educational Psychology, 1975 (Feb), Vol. 45(1), 73-76. Journal Abstract.

46. Wright, Anthony A. (U. Texas, Graduate School of Biomedical Sciences, Sensory Sciences Center, Houston) *Psychometric and psychophysical theory within a framework of response bias.* Psychological review, 1974 (Jul), Vol. 81(4), 322-347. Journal Abstract.

APPENDIX B. DISTRIBUTION LIST

TECOM Project No. 7 CO IL7 TT1 001

A New Approach Toward Obtaining Quantifed Subjective Test Data

Addressee	Final Report
Commander US Army Test and Evaluation Command ATTN: DRSTE-ME Aberdeen Proving Ground, MD 21005	3
Commander US Army Materiel Development and Readiness Command 5001 Eisenhower Avenue Alexandria, VA 22333	1
HQDA (DAMA-ARZ-B) Washington, DC 20314	1
HQDA (DAEN-ZA) Forrestal Bldg Washington, DC 20314	1
Assistant Secretary of Army (R&D) ATTN: Dr. Emerson Washington, DC 20310	1
Associate for Documentation and Accession Psychological Abstracts American Psychological Association 1200 Seventeenth Street, NW Washington, DC 20036	1
Institute for Applied Technology National Bureau of Standards Washington, DC 20234	1
Director The Army Library The Pentagon, Rm 1A526 Washington, DC 20310	2
Director Naval Research Laboratory ATTN: Code 2627 Washington, DC 20375	1
Smithsonian Institution Office of the Assistant Secretary for Science (SI 114) 1000 Jefferson Drive, SW Washington, DC 20560	1
Commander US Army Electronic Proving Ground ATTN: STEEP-MT-I Fort Huachuca, AZ 85613	1
Commander US Army Yuma Proving Ground ATTN: STEYP-MMI Yuma, AZ 85364	1
Commander Naval Training Equipment Center ATTN: DRCPM-TND Orlando, FL 32813	1
President US Army Infantry Board ATTN: Tech Director Fort Benning, GA 31905	1
Commander US Army Armament Materiel Readiness Command ATTN: DRSAR-RDM Rock Island, IL 61201	1
President US Army Armor and Engineer Board ATTN: Tech Director Fort Knox, KY 40121	1
Commander US Army Natick R&D Command ATTN: DRDNA-E DRDNA-UBS Natick, MA 01760	 5 1
Commander US Army Research Institute of Environmental Medicine ATTN: SGRD-VE-ADJ Natick, MA 01760	1
Director US Army Human Engineering Laboratory ATTN: DRXHE-D DRXHE-PC Aberdeen Proving Ground, MD 21005	 1 1
Director US Army Materiel Systems Analysis Activity ATTN: DRXSY-D DRXSY-HEL Aberdeen Proving Ground, MD 21005	 2 3

Addressee	Final Report
President US Army Airborne, Communications and Electronics Board ATTN: STEBG-TD Fort Bragg, NC 28307	 1
Commander US Army Research Office ATTN: Information Processing Office PO Box 12211 Research Triangle Park, NC 27709	 1
Commander US Army Armament R&D Command Picatinny Arsenal ATTN: DRDAR-DRG DRDAR-M110E2 Dover, NJ 07801	 1 1
Commander US Air Force Test and Evaluation Center ATTN: AFTEC/XR Kirtland Air Force Base, NM 87115	 1
Commander US Army White Sands Missile Range ATTN: STEWS-TE White Sands, NM 88002	 1
President US Army Field Artillery Board ATTN: ATZR-BDTD Fort Sill, OK 73503	 1
President US Army Air Defense Board ATTN: ATZC-D-MT Fort Bliss, TX 79916	 1
Commander TRADOC Combined Arms Test Activity ATTN: TCATA-OP Fort Hood, TX 76544	 1
Commander US Army Dugway Proving Ground ATTN: STEDP-MT-L STEDP-SC Dugway, UT 84022	 1 1
Administrator Defense Documentation Center ATTN: DDC-T Alexandria, VA 22314	 2
Chief of Naval Research 800 N. Quincy Street Arlington, VA 22309	1

Addressee	Final Report
Director US Army Research Institute for the Behavioral and Social Sciences Reference Service 1300 Wilson Blvd Arlington, VA 22209	1
Commander US Army Operational Test and Evaluation Agency ATTN: DACS-TEO-N 5600 Columbia Pike Falls Church, VA 22041	 1
Commander US Army Tropic Test Center ATTN: STETC-TD STETC-TA STETC-TD-T STETC-TD-O STETC-TD (Tech Lib) STETC-MO (Ed Rm) Fort Clayton, CZ	 10 1 1 1 2 2